Phys. Educ. **21** 1986. Printed in Great Britain

The paradox of Schrödinger's cat

C N Villars

Erwin Schrödinger first described the thought-experiment which has since become known as 'the paradox of Schrödinger's cat' 51 years ago (Schrödinger 1935, account in Jammer 1974 p 217). In recent years, popular accounts of quantum mechanics have tended to adopt one or other of the philosophically most extreme solutions to this paradox, i.e. the consciousness hypothesis or the many worlds interpretation. The present essay is an attempt to redress the balance by describing what the present author takes to be the orthodox solution to the paradox, which explains the paradox, without recourse to such counterintuitive notions as a cat simultaneously dead and alive or a universe continually splitting into multiple worlds, as being due to a misapplication of the quantum formalism.

It is doubtful whether many physicists accept either the consciousness hypothesis or the many worlds interpretation, though, as far as the author knows, there has never been a survey of physicists' opinions on these matters. However, whatever the general opinion, there do exist numerous more or less identical and generally accepted statements of the fundamental rules, or axioms, of quantum mechanics from which what might be called an 'orthodox' solution to the paradox can be derived.

Schrödinger's thought-experiment

Let us first consider Schrödinger's thought-experiment. A cat is enclosed in a sealed steel chamber so as to be completely isolated from the outside world. Along with the cat in the chamber is a device capable of releasing a sufficient quantity of a deadly poison into the atmosphere of the chamber to kill the cat instantly. The operation of this device is triggered by a quantum-mechanical measurement process.

In his original account of the experiment, Schrödinger imagined the triggering measurement process to be an observation of the radioactive decay of a small quantity of radioactive material. However, as Richard Schlegel (1980) points out, the quantum-mechanical explanation of radioactive decay does not usually involve representing the radioactive nucleus in a superposition of 'particle emitted' and 'particle not emitted' states. Hence, Schrödinger's original example does not readily lead to the kind of superposition that is needed for the 'paradox' to be derived. Consequently, subsequent accounts have often substituted alternative processes. In the present account, the triggering process will be assumed to be an observation of the spin component of an electron using a Stern–Gerlach apparatus.

The experiment is shown schematically in figure 1. A beam of electrons, all initially prepared in the same spin state (corresponding to a spin vector pointing vertically out of the page in figure 1), is split in two by an inhomogeneous magnetic field. The resulting 'upper' beam corresponds to states described by spin vectors pointing up the page, and the 'lower' beam to states described by spin vectors pointing down the page. Geiger tubes, or any other instruments capable of detecting single electrons, are placed in the path of each beam at points 1 and 2. The detectors are connected to the poison-releasing device in such a way that detection of an electron at 1 triggers the device and kills the cat, whilst detection at 2 does not trigger the device and the cat lives.

The paradox

Suppose that the experiment is performed with a very low intensity electron beam, so that, in fact, only *one* electron passes through the apparatus during the experiment. The electron wave representing this single electron is split into two components by the magnetic field, so that the state of the electron when it reaches the detectors is a superposition of the 'up' and 'down' spin states. Thus,

$$\psi_e=(\sigma_{\text{up}}+\sigma_{\text{down}})/\sqrt{2}, \qquad (1)$$

where σ_{up}, σ_{down} represent the 'up' and 'down' spin states respectively.

The electron next interacts with the detectors at 1 and 2. As a result of this interaction, the state of the combined system of electron+detectors becomes

$$\psi_{e+D}=(\sigma_{\text{up}}D_1+\sigma_{\text{down}}D_2)/\sqrt{2}. \qquad (2)$$

Christopher N Villars *is a systems analyst in the computer division of a public company. He obtained a BA in philosophy from the University of Birmingham in 1975. Until 1980 he worked as a physics teacher in secondary schools. His main interest lies in the philosophical problems of modern physics. His publications include 'Observables, states and measurements in quantum physics' 1984* Eur. J. Phys. **5** *177–83.*

0031-9120/86/040232+06$2.50

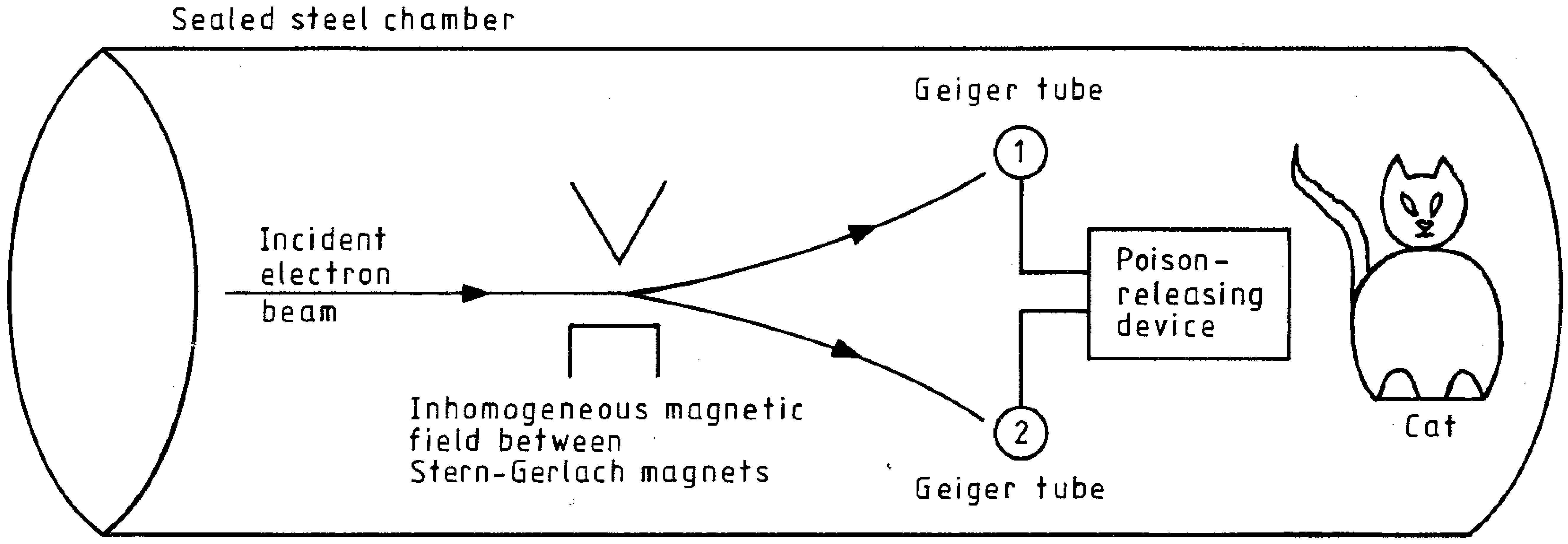

Figure 1 Schematic representation of the Schrödinger cat thought-experiment

Here, D_1 and D_2 are states corresponding to the electron being detected at 1 or 2 respectively. The equal coefficients, $1/\sqrt{2}$, show that there is an equal probability, $(1/\sqrt{2})^2=\frac{1}{2}$, of the electron being detected at 1 or 2.

Finally, as a result of the further interaction between the detectors and the cat, *via* the poison-releasing device, the state of the total system of electron+detectors+cat becomes

$$\psi_{\mathrm{e+D+C}}=(\sigma_{\mathrm{up}}D_1C_{\mathrm{dead}}+\sigma_{\mathrm{down}}D_2C_{\mathrm{alive}})1/\sqrt{2}, \quad (3)$$

where C_{dead} and C_{alive} are states corresponding to the cat being dead or alive respectively.

Thus, according to the laws of quantum mechanics, at the end of this experiment the state of the apparatus in the steel chamber is a superposition of states in which the 'dead' and 'alive' cat states are, as Schrödinger described it, 'mixed or smeared together by equal amounts' (Jammer 1974 p 217). The cat is neither definitely dead nor definitely alive, but, in some sense, both dead and alive at the same time. Yet, it is certain that if the chamber were opened the cat would be found to be dead or alive, and not in any kind of intermediate condition. This discrepancy between the superposition apparently predicted by quantum mechanics and the definite actual results of the experiment is the paradox of Schrödinger's cat.

It is worth noting that this paradox is merely an exaggerated version (a 'burlesk', Schrödinger called it (Schlegel 1980 p 176)) of the so called 'measurement problem' in quantum mechanics. Thus there is really no need to introduce the cat at all. The paradox already exists after the interaction between the electron and the detectors at 1 and 2. Equation (2) implies that after this interaction the apparatus is in a superposition of states corresponding, in some sense, to both detectors being activated. Yet, on inspection, either detector 1 or detector 2, and never both, would be found to have been activated.

The transition from a superposition of states, such as that represented by equation (3), to one of the terms of the superposition, corresponding to the definite occurrence of an experimental result, is called the 'collapse' or 'reduction' of the state vector during a measurement. On opening the steel chamber the cat is found to be either dead or alive, implying that the superposition (3) has been reduced to one of its terms, i.e. either

$$\psi_{\mathrm{e+D+C}} \xrightarrow[\text{reduced to}]{} \sigma_{\mathrm{up}}D_1C_{\mathrm{dead}} \quad (4a)$$

or

$$\psi_{\mathrm{e+D+C}} \xrightarrow[\text{reduced to}]{} \sigma_{\mathrm{down}}D_2C_{\mathrm{alive}}. \quad (4b)$$

In order to account for the reduction of state vectors during a measurement, and hence also to explain the paradox of Schrödinger's cat, some physicists have introduced additional hypotheses. The two most widely discussed of these are the consciousness hypothesis and the many worlds interpretation.

The consciousness hypothesis

When the steel chamber is opened and the experimenter looks inside, the cat is found to be either dead or alive. Immediately before this, according to the laws of quantum mechanics, the cat is neither dead nor alive, but in a superposition of the 'dead' and 'alive' states. This has led some authors to suggest that it is the act of conscious observation of the cat by the experimenter that reduces the superposition and determines the cat to be either dead or alive.

This possibility was first hinted at by John von Neumann (1932, account in Jammer 1974 ch 11) and subsequently developed explicitly by Fritz London and Edmond Bauer (1939, account in Jammer 1974 ch 11) and Eugene Wigner (1967). According to this point of view, until it is consciously perceived by the experimenter the cat exists in an indeterminate condition, in some sense, half dead and half alive. It is the experimenter's consciousness, at the moment when the chamber is opened and the experimenter looks at the cat, which determines the cat to be definitely dead or alive.

Clearly, this 'consciousness hypothesis' is strongly counterintuitive. Few people ordinarily think that their perceptions create what they see, at least, not to the extent that is suggested here. Perception is generally thought of as a conscious registering of things already there. A table unperceived in another room, or Schrödinger's cat unperceived in its steel chamber, are still considered to exist, and to exist in a definite state, even though they are unperceived.

Many worlds

A very different solution to the quantum measurement problem, and hence to Schrödinger's cat paradox, was proposed by Hugh Everett III in 1957. The view was developed by Bryce DeWitt and others (1973) and has been adopted in several recent popular accounts of quantum mechanics (Davies 1980, Gribbin 1984).

Everett's view avoids the problem of explaining the reduction of state vectors during measurements by assuming that *it never happens*. Instead, it is supposed that the universe splits into two or more separate 'worlds', in each of which one of the possible experimental results is actualised. Thus, the final state of the apparatus in the steel chamber in Schrödinger's cat experiment is assumed to be accurately described by the superposition represented by equation (3). However, this is interpreted, not as representing the cat in the unimaginable condition of being simultaneously dead and alive, but rather as describing two separate worlds. In one of these the cat is dead, whilst in the other it is alive.

During the experiment, not only the electron, the detectors and the cat, but also the experimenter split into two worlds so that when, in one world, the experimenter opens the chamber and finds the cat dead, in the 'other' world, the 'other' experimenter opens the chamber and finds the cat alive. Thus, the discrepancy between the definite experimental result and the superposition predicted by quantum mechanics is resolved by supposing the universe to be continually branching into multiple worlds, all completely isolated from one another.

Like the consciousness hypothesis, the many worlds interpretation is strongly counterintuitive. Most people think of themselves as *one* self existing in *one* world and find it difficult to grasp the idea of an infinity of actual, noninteracting worlds, in each of which a different version of their self realises a different possible life. In fact, both the consciousness hypothesis and the many worlds interpretation are so contrary to common sense that there would need to be a very good reason why they should be given any serious consideration at all. To advocates of these points of view, this reason is the 'strangeness' of the orthodox view of the reduction of state vectors during measurements.

An orthodox interpretation

The fundamental rules, or axioms, of quantum mechanics distinguish two fundamentally different kinds of physical process. These are *ordinary physical interactions* and *measurements*. Ordinary physical interactions, for example, the interaction between an electron and an electric or magnetic field or between one electron and another, are completely described by the law of continuous evolution of states. This is given in Postulate 5 of the introductory account of the quantum formalism by Daniel Gillespie (1970) and Rule 3 in the more rigorous treatment by Bernard d'Espagnat (1976). (These two examples are taken to be representative. Equivalent postulates can be found in all other accounts.)

A *measurement* is an interaction with an observing instrument (taken to be of macroscopically large size) leading to the production of a particular observational result, for example, the interaction between a photon and a photoemulsion in an interference experiment or between an electron and the Geiger detectors in a Stern–Gerlach experiment. Measurements involve, in addition to continuous evolution of states, discontinuous changes of state governed by a probabilistic law. This law is given in Postulates 3 and 4 (Gillespie) and Rules 4–10 (d'Espagnat). This additional law, unique to measurements, describes the reduction of state vectors during a measurement. Thus, according to the orthodox quantum formalism, the reduction of state vectors during a measurement *is one of the axioms of the theory*.

It follows that, from the standpoint of the orthodox formalism, *the paradox of Schrödinger's cat arises from a misapplication of the quantum rules*. Specifically, it is incorrect to claim that the state of the combined system of electron+detectors after their interaction is given by equation (2) above. This interaction is not an ordinary physical interaction, but a *measurement*, and, hence, involves a reduction of the state vector, i.e. either

$$\psi_{\mathrm{e+D}} \xrightarrow{\text{reduced to}} \sigma_{\mathrm{up}} D_1$$

or

$$\psi_{\mathrm{e+D}} \xrightarrow[\text{reduced to}]{} \sigma_{\mathrm{down}} D_2 \ . \qquad (5)$$

Thus, the state of the system of electron+detectors after their interaction is not the superposition given by equation (2), but is reduced to one of the terms on the right-hand sides of equations (5). As a result of the further interaction, *via* the poison-releasing device, with the cat, the final state of the total system of electron+detectors+cat becomes either

$$\sigma_{\mathrm{up}} D_1 C_{\mathrm{dead}} \quad \text{or} \quad \sigma_{\mathrm{down}} D_2 C_{\mathrm{alive}}$$

and not the superposition given by equation (3).

In this way, the account of the experiment given by orthodox quantum mechanics is completely in accord with what common sense expects: the electron wave interacts with the detectors and activates one of them. Depending on which detector is activated, poison either is or is not released and the cat is either dead or alive, regardless of whether or not the chamber is opened and the result consciously perceived by the experimenter.

In essence, the orthodox interpretation of the reduction of state vectors during measurements proposes that we simply accept the dualism of two kinds of physical process in quantum physics. The postulate of the reduction of state vectors during measurements is to be regarded as an axiom of equal status to the postulate of the law of continuous evolution of states during ordinary physical interactions.

Discussion

To advocates of the consciousness hypothesis or the many worlds interpretation, the postulate, in orthodox quantum mechanics, of the reduction of state vectors during a measurement is apparently so 'strange' (Wigner 1967 p 155), or 'bizarre'(Gribbin 1984 p 173), that they reject it completely. They then have to introduce additional hypotheses (consciousness or many worlds) to explain what *it* explained, i.e. how it is that a quantum measurement process leads to a definite result and not a superposition. However, these additional hypotheses would generally be regarded as even more strongly counterintuitive than the orthodox postulate they replace.

Why do the reductions of state vectors during quantum measurements seem 'strange' to some physicists? The reasons most frequently given are: firstly, these discontinuous events are unlike anything that happens in classical physics, where waves always evolve in a continuous way; secondly, they only occur during measurements and never during ordinary physical interactions; and thirdly, the orthodox theory offers no explanation of *how* these events occur, they are simply postulated in the axioms of the theory. However, if the novel physical meaning of observables, states and measurements in quantum physics is taken into account, these 'strange' aspects of the reduction postulate can become quite acceptable intuitively.

In classical physics, all physical objects were completely described in terms of intrinsic properties (i.e. properties they possessed in themselves independently of whether or not they were observed) and laws that specified how these properties changed when the objects interacted with one another. Measurement was conceived as the passive registration of these intrinsic properties. Such a view was possible because the classical laws allowed in principle that the effect of the observation interaction on the observational result could always be made negligibly small or compensated for by further measurements. Thus, classical measurements did not contribute significantly to what was observed and, as a result, the process of measurement itself could be ignored. Consequently, measurements played no special role in classical physics. They were just ordinary physical interactions, describable by the same laws as any other physical interaction.

The fundamental discovery of quantum physics, namely the discovery of the quantisation of physical action, undermines this classical conception of measurements. As Niels Bohr, in particular, has repeatedly emphasised, in quantum physics, due to the finite and indivisible nature of the quantum of action, the interaction between the observed object and the observing instrument forms an inseparable part of what is observed (Bohr 1963 p4). Consequently, in quantum physics, the observational result must be taken to represent a property of the observation interaction as a whole, not an intrinsic property of the observed object alone. Thus, quantum observables are *interactional properties*, i.e. properties of interactions with observing instruments, not intrinsic properties.

The state of an object in classical physics represented the totality of the actual values of the intrinsic properties of the object at any given time. Since quantum observables are not intrinsic properties, it follows that states must also have a different meaning in quantum physics. In fact, a quantum state represents the *potentiality* of the object for interacting with observing instruments to produce particular observational results. Any quantum state function, $\psi(t)$, can be expressed as a linear superposition of eigenfunctions of any arbitrary quantum observable, A. Thus,

$$\psi(t)=c_1\psi^A_1(t)+c_2\psi^A_2(t)+\ldots=\sum_i c_i\psi^A_i(t) \quad , \qquad (6)$$

where the coefficients, c_i, are complex numbers and the $\psi^A{}_i(t)$ are the eigenfunctions of A. Each of these eigenfunctions represents one possible state that the observed object may be left in after the measurement of A. Thus, each eigenfunction corresponds to one *possible* observation interaction, and the complete set of eigenfunctions represents the *range of possible observation interactions* open to the object before the measurement of A. Formally, the square moduli of the coefficients, *viz*$|c_i|^2$, represent the relative probability of each possible observation interaction, or, interpreted physically, the *tendency of the object towards each possibility*. These two factors, namely range of possible interactions and tendency towards each possibility, together constitute the *potentiality* of the object for interacting with observing instruments to produce particular observational results. Thus, a quantum state function does not represent an *actual* state of affairs, but rather it represents the *potentiality* of the object.

Clearly, due to the quantisation of action, observables and states have a completely different meaning in quantum physics from that which they had in classical physics. If this point is not recognised, and observables and states are interpreted in classical terms, the reduction of a state vector during a measurement seems to represent the instantaneous collapse, or sudden concentration, of a widely extended wave of actual physical disturbance into some particular small region, with the simultaneous disappearance of the disturbance from all other regions. Such events would be unlike anything that occurs in classical physics and consequently seem strange. However, if the novel meaning of quantum observables and states is taken into account, the reduction of a state vector during a measurement acquires a completely different meaning.

The state vector before it is reduced represents the potential observation interactions of the object, with the intensity of the function at any point giving the probability of an observation interaction producing a result at that point. Thus, before the measurement, the object has a range of potential observation interactions open to it represented by a superposition of eigenstates, as shown, for example, in equations (1) and (6). During the measurement, the object actualises one of these potential interactions and at the same time excludes the others. This corresponds to the reduction of the state vector represented, for example, by equations (5). After the measurement, the object is left in the eigenstate corresponding to the observation interaction that actually occurred. The object no longer has a range of potential observation interactions (of the kind just measured) open to it. In actualising one possibility, it excludes the others, just as in selecting one possible route at a road junction we immediately cut ourselves off from the possibility of selecting any of the others. Thus, the reduction of the state vector during a measurement represents *a logical reduction of the range of potential observation interactions open to the observed object due to its actualisation of one of them*, and not the mysterious collapse of a wave of actual disturbance in the classical sense.

Why do reductions of state vectors only occur during measurements and not during all physical interactions generally? This question arises because classical physics makes no distinction between measurements and ordinary physical interactions. Classical states change in essentially the same way regardless of whether the object is interacting with observing apparatus or with any other physical object. Thus, from a classical point of view, it seems strange that the state of a microphysical object changes in a unique way during observation interactions.

However, as indicated above, the term 'state' does not have the same meaning in quantum physics as in classical physics. The state of a microphysical object represents its potential observation interactions. The discontinuous reduction of this state during a measurement represents the actual occurrence of one of these potential interactions. When a microphysical object interacts with an external field of force, or with another object other than an observing instrument, though its potentiality for observation interactions (i.e. its state) will generally change, it cannot actualise any of these potential interactions. By definition, observation interactions can only actually occur during interactions with observing instruments. Hence, reductions of state vectors only occur during observation interactions because, given the physical meaning of these events, it is only during such interactions that they can occur.

How do reductions of state vectors during measurements occur? Orthodox quantum mechanics provides no physical explanation, in the sense of a derivation from more general physical laws, of these events. They are simply postulated in the axioms of the theory. This, in itself, is not mysterious. No theory can be expected to explain, in the above sense, what it postulates in its axioms. Nevertheless, some physicists consider the reduction postulate to be *ad hoc* and in need of further explanation. Intuitively, they argue, observing instruments are not essentially different from any other physical object. Why, then, should observation interactions not obey the same law of continuous evolution of states that applies to ordinary physical interactions?

However, the lack of a physical explanation of the reduction postulate does not indicate the incompleteness of quantum theory but, rather, is an unavoidable consequence of the quantisation of action. As indicated above, the existence of an indivisible quantum of action implies that a quantum measurement process cannot be completely analysed in the classical sense and, hence, cannot be completely explained by the theory. Thus, the inability to explain the reduction postulate in orthodox quantum mechanics is a consequence of the indivisibility of the quantum of action. As a result of this inability to completely explain observation interactions, they become the irreducible data in terms of which the theory is formulated. Quantum states are defined in terms of potential observation interactions and these interactions are treated in separate axioms of the theory. Hence, the quantum formalism makes a 'logical distinction' (Bohr 1963 p 6) between observing instruments and ordinary physical objects and simply does not support the intuitive expectation, based on the experience of classical physics, that observing instruments are not essentially different from any other physical object. In quantum physics, observing instruments define the framework within which physical reality is described and cannot themselves be completely explained within that framework.

The point of view expressed above, derived mainly from Bohr, is controversial. In particular, some physicists continue to search for a physical explanation of the reduction of state vectors during measurements. To date, these attempts have not been successful (for a review of various approaches, see d'Espagnat 1976 chs 16 and 17). Consequently, at present, it remains an open question whether any such explanation will ever be possible. However, from the point of view of the orthodox interpretation, failure seems inevitable. The essential lesson of quantum physics seems to be that there is a fundamental limitation on our ability to analyse a measurement process due to the existence of a finite and indivisible quantum of action. The 'strangeness' of the orthodox view of quantum measurements seems to be due to failure to recognise the novel physical meaning of observables, states and measurements in quantum physics which this limitation implies. If these novel meanings are taken into account, this 'strangeness' disappears and the orthodox interpretation of quantum measurements, and hence also of Schrödinger's cat paradox, seems quite satisfactory.

Conclusion

It is not necessary to resort to the idea of a direct influence of consciousness or a multiple splitting of the universe in order to explain the paradox of Schrödinger's cat. A reduction of the state vector, leading to a definite experimental result, can be assumed to occur quite objectively in the measurement process, so that the cat is either dead or alive inside the steel chamber before it is opened by the experimenter.

References

Bohr N 1963 'Quantum Physics and Philosophy' in *Essays 1958–1962 on Atomic Physics and Human Knowledge* (New York: Wiley-Interscience)

Davies P 1980 *Otherworlds* (London: Dent)

DeWitt B S and Graham N 1973 *The Many-Worlds Interpretation of Quantum Mechanics* (New Jersey: Princeton University Press)

d'Espagnat B 1976 *Conceptual Foundations of Quantum Mechanics* (Reading, MA: Benjamin)

Everett H III 1957 '"Relative state" formulation of quantum mechanics' *Rev. Mod. Phys.* **29** 454–62

Gillespie D T 1970 *A Quantum Mechanics Primer* (New York: International Textbook Company)

Gribbin J 1984 *In Search of Schrödinger's Cat* (London: Wildwood House)

Jammer M 1974 *The Philosophy of Quantum Mechanics* (New York: Wiley-Interscience)

London F and Bauer E 1939 *La Theorie de l'Observation en Mechanique Quantique* (Paris: Hermann)

von Neumann J 1932 *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer)

Schlegel R 1980 *Superposition and Interaction* (Chicago: University of Chicago Press)

Schrödinger E 1935 'Die gegenwartige Situation in der Quantenmechanik' *Die Naturwissenschaften* **23** 807–12

Wigner E P 1967 'The problem of measurement', 'Remarks on the mind-body problem' and 'Two kinds of reality' in *Symmetries and Reflections* (Indiana: Indiana University Press)

Science, Technology and Drama Competition

Following a successful first year, the British Association's Science, Technology and Drama Competition will again take place in 1986/7. Primary and secondary school groups are asked to devise short theatrical presentations in three categories. These are an event from the history of science, technology or engineering; a socially relevant issue arising out of science and its applications; or an illustration of a scientific concept.

The British Association believes that this type of activity helps to get its aims (promoting greater understanding of science and technology) over in the most interesting and fun way. Also, just as importantly, it should bring science into other lessons by encouraging cross-curriculum links.

The Marshland Middle School from Thorne, near Doncaster, won the 1985/6 finals with their play *Thermos Factor*, which illustrated various methods for the transfer of heat. The finals took place late spring at the Sherman Theatre, Cardiff, and the winners received £200 and a Lloyds Bank trophy. Lloyds Bank has agreed to sponsor the competition again as part of their support for young people's education and training.

Schools from the Middlesex and Surrey region will be joining those from Humberside, Lincolnshire, Somerset and Wiltshire, South Wales and Yorkshire and North Derbyshire in this year's competition.
It is hoped that schools from another county in the South West and from parts of Scotland will also be entering.

For more information, contact Dr Peter Briggs, British Association, Fortress House, 23 Savile Row, London W1X 1AB (tel. 01-734 6010).